

**MAE116 – Noções de Estatística**  
**Grupo A - 1º semestre de 2012 - GABARITO**  
 Lista de exercícios 2 – Estatística Descritiva – C L A S S E

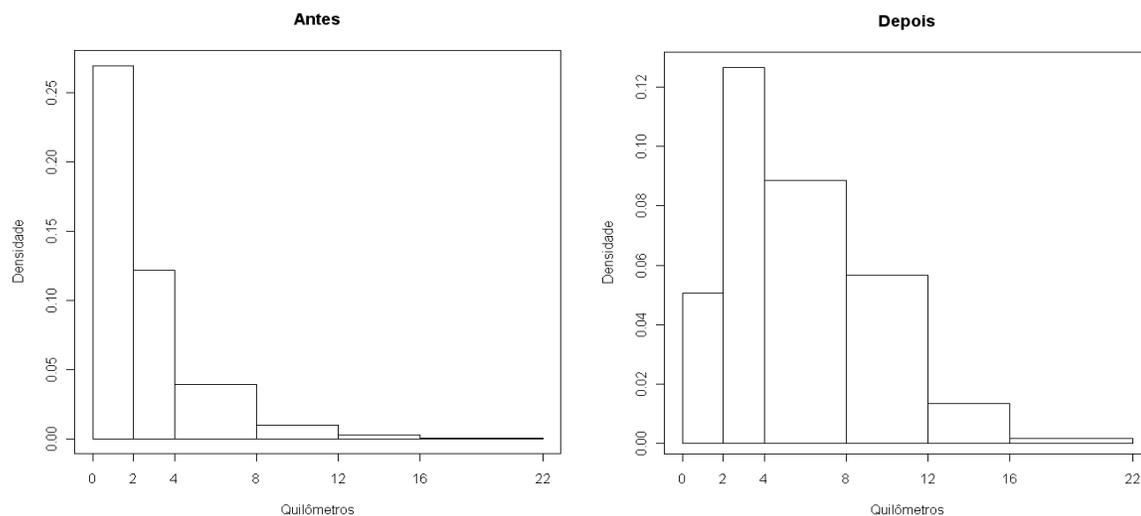
**Exercício 01**

	Antes do treinamento			Depois do treinamento			
	base	freq. absoluta	freq. relativa (%)	densidade de freq.	freq. absoluta	freq relativa (%)	densidade de freq.
0  -- 2	2	442	53,97	26,98	80	10,13	5,06
2  -- 4	2	200	24,42	12,21	200	25,32	12,66
4  -- 8	4	130	15,87	3,97	280	35,44	8,86
8  -- 12	4	34	4,15	1,04	179	22,66	5,66
12  -- 16	4	10	1,22	0,31	43	5,44	1,36
16  -- 22	6	3	0,37	0,06	8	1,01	0,17
total		819	100,00		790	100,00	

(a) Qual é a variável em estudo?

A variável em estudo corresponde ao **número de quilômetros corridos sem parar**.

(b) Construa 2 histogramas, um para o teste antes do treinamento e outro para depois. Compare.



**OBSERVAÇÕES** (da Beti sobre a construção do histograma): A ideia é fazer este exercício à mão, na lousa. As densidades de frequência nas tabelas estão multiplicadas por 100, para facilitar a construção do gráfico. As escalas dos histogramas são diferentes, pois é saída de computador (não sei qual pacote).

Os possíveis comentários sobre as tabelas e histogramas são:

**MAE116 – Noções de Estatística**  
**Grupo A - 1º semestre de 2012 - GABARITO**  
**Lista de exercícios 2 – Estatística Descritiva – C L A S S E**

- 1) Observando os histogramas acima, pode-se notar que, após o treinamento, os alunos conseguiram correr mais, dado que a faixa de maior frequência relativa (53,97 %) de ANTES foi [0,2) km, enquanto que DEPOIS do treinamento a faixa de maior frequência relativa (35,44%) foi [2,4) km.
- 2) Além disso, apenas 5,74% (= 4,15+1,22+0,37) de ANTES conseguiram correr mais de 8km, enquanto que DEPOIS 29,11% (= 22,66+ 5,44+ 1,01) conseguiram correr mais de 8km.
- 3) Observa-se uma forte assimetria apresentada no histograma para os dados ANTES do treinamento, pois as frequências estão concentradas no lado esquerdo do gráfico: 78,39% (=53,97+24,42) nas de até 4 km. No histograma DEPOIS do treinamento o qual os alunos foram submetidos, essa assimetria diminuiu.

**Exercício 02** - Os dados da tabela estão no arquivo: **DadosA2012ClasseEx2.csv**

**(a)** Classifique as variáveis listadas.

**Seção:** qualitativa nominal

**Nota em Direito:** quantitativa contínua, eventualmente discretizada

**Nota em Política:** quantitativa contínua

**Nota em Estatística:** quantitativa contínua, eventualmente discretizada

**Conceito em Inglês:** qualitativa ordinal

**Conceito em Metodologia:** qualitativa ordinal

**(b)** Calcule média, moda, mediana e desvio padrão das variáveis notas dos cursos de Direito, Política e Estatística.

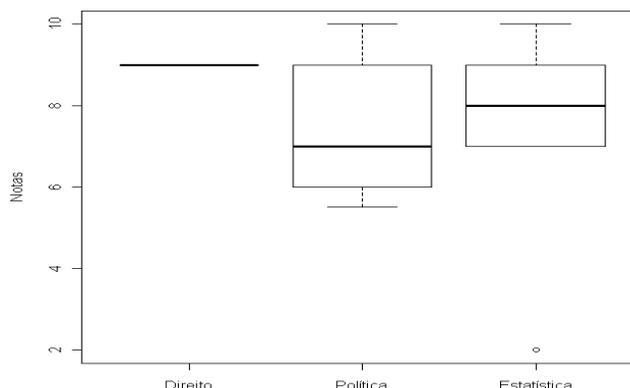
Variável	Média	Moda	Mediana	Desvio padrão
Direito	9,00	9	9	0,00
Política	7,30	9	7	1,54
Estatística	7,96	8	8	1,39

Comentários sobre as medidas resumo:

- Pela tabela acima, pode-se perceber que a nota em **Direito** é uma variável que assume valor constante igual a 9, não tendo nenhuma variabilidade (desvio-padrão igual a zero).
- A nota em **Política** possui uma variabilidade levemente maior que a de **Estatística**, dado que apresentou o desvio padrão igual a 1,54, comparado com 1,39 de Estatística.
- A moda em Política é maior, mas tanto a nota mediana, quanto a nota média em Estatística são maiores que em Política.

**MAE116 – Noções de Estatística**  
**Grupo A - 1º semestre de 2012 - GABARITO**  
**Lista de exercícios 2 – Estatística Descritiva – C L A S S E**

- (c)** Compare e indique as diferenças existentes entre as distribuições das variáveis notas dos cursos de Direito, Política e Estatística, utilizando o “boxplot” e as medidas calculadas em



**(b).**

Os possíveis comentários sobre a tabela do item (b) e o *boxplot* são:

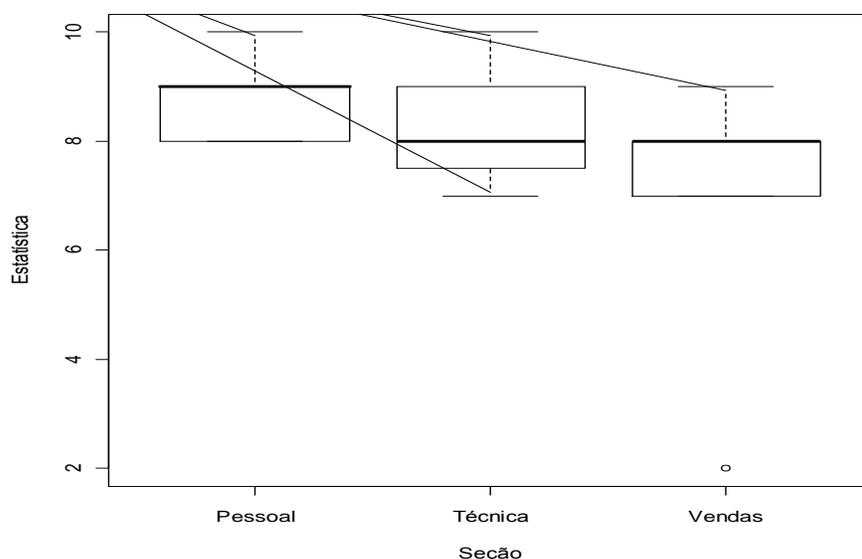
- 1) O *boxplot* das notas por disciplina reforça os comentários feitos no item (b): a nota em Direito é constante, a nota em Política possui, de fato, maior variabilidade (a caixa do *boxplot* é maior) comparada com a de Estatística.
- 2) Note que, pelo desvio padrão essa variabilidade era **levemente** maior, mas pelo *boxplot*, percebe-se que as variabilidades são bem diferentes.
- 3) Apesar da moda ser maior em Política, a nota média e a nota mediana em Estatística são maiores.
- 4) Na verdade 75% das notas em Estatística são maiores que 50% das notas em Política.
- 5) Pode-se observar a presença de um valor extremo na variável Nota em Estatística (alguém teve nota 2). Isso pode explicar a razão do desvio padrão ser próximo ao desvio padrão das notas em Política, embora, os gráficos expressem uma diferença grande na variabilidade.

- (d)** Compare o aproveitamento dos funcionários na disciplina **Estatística** segundo a seção a qual eles pertencem.

A tabela abaixo apresenta o aproveitamento dos funcionários na disciplina **Estatística** de acordo com a seção a qual eles pertencem. E, em seguida, tem-se o *boxplot* da nota em Estatística por seção.

<b>Seção</b>	<b>n</b>	<b>Média</b>	<b>Mediana</b>	<b>Desvio-padrão</b>	<b>Coefficiente de Variação</b>
Pessoal	7	8,71	9	0,76	8,67 %
Técnica	7	8,28	8	1,11	13,43 %
Vendas	11	7,27	8	1,90	26,15 %

**MAE116 – Noções de Estatística**  
**Grupo A - 1º semestre de 2012 - GABARITO**  
 Lista de exercícios 2 – Estatística Descritiva – C L A S S E



Possíveis comentários:

- 1) Nota-se que os funcionários da seção **Pessoal** obtiveram uma nota média (e mediana) superior a nota média (e mediana, respectivamente) dos funcionários das demais seções. O boxplot fornece uma representação visual das diferenças das medianas e dos quartis, de modo que 50% dos funcionários da seção **Pessoal** obtiveram nota em Estatística maior que 75% dos funcionários Técnicos e de 100% dos de **Vendas**.
- 2) Quanto à dispersão (ou variabilidade) das notas de Estatística, tem-se que na seção **Pessoal** apresentou o menor desvio padrão, seguido de **Vendas**. Este indício de menor variabilidade por parte das notas em Estatística dos funcionários da seção **Pessoal** pode ser reforçado pelo coeficiente de variação (que mede a dispersão dos dados relativamente à média), que é o menor dos três apresentados na tabela.
- 3) No entanto, pelo *boxplot*, as dispersões de **Pessoal** e **Vendas** são parecidas, exceto que na seção **Vendas** há um valor extremo (nota igual a 2).
- 4) Tanto pelas medidas, como pelos gráficos, conclui-se que os funcionários da seção **Pessoal** tiveram um melhor aproveitamento em Estatística.

- (e) Construa uma tabela de contingência para as variáveis **Seção** e **Inglês**. Determine as frequências relativas e compare o desempenho dos funcionários em inglês segundo a seção a qual eles pertencem.

Segue a tabela de contingência para as variáveis **Seção** e **Inglês**.

Seção	Inglês				Total
	A	B	C	D	
Pessoal	1	3	0	3	7
Técnica	1	4	1	1	7
Vendas	2	1	5	3	11
<b>Total</b>	4	8	6	7	25

**MAE116 – Noções de Estatística**  
**Grupo A - 1º semestre de 2012 - GABARITO**  
**Lista de exercícios 2 – Estatística Descritiva – C L A S S E**

---

A tabela abaixo apresenta as **frequências relativas por total das linhas**. O desempenho dos funcionários varia de acordo com a seção que pertencem. Os funcionários da seção Pessoal obtiveram, em sua grande maioria, conceitos B e D, ambos com a mesma frequência relativa (42,86%). A maioria dos funcionários da seção Técnica obteve conceito B (57,14%). Já os funcionários da seção de Vendas apresentaram maior percentual no conceito C (45,45%). Na última linha da tabela estão apresentadas as frequências relativas dos conceitos obtidos por todos os funcionários. Note que a maioria (32%) obteve conceito B, este sendo o maior obtido para os funcionários das seções Pessoal e Técnica, porém não para os funcionários da seção Vendas. Portanto, pode-se concluir que os funcionários da seção Técnica apresentaram o melhor desempenho em inglês, seguido dos funcionários da seção Pessoal.

Seção	Inglês				Total
	A	B	C	D	
Pessoal	14,29 %	42,86 %	0,00 %	42,86 %	100,00 %
Técnica	14,29 %	57,14 %	14,29 %	14,29 %	100,00 %
Vendas	18,18 %	9,09 %	45,45 %	27,27 %	100,00 %
<b>Total</b>	16,00 %	32,00 %	24,00 %	28,00 %	100,00 %

### Exercício 3

Algumas variáveis registradas no estudo são:

(a) Classificação:

**Sexo:** qualitativa

**Idade:** quantitativa contínua

**Unidade da USP:** qualitativa

**Estado Geral de Saúde antes (EGS\_a):** quantitativa contínua

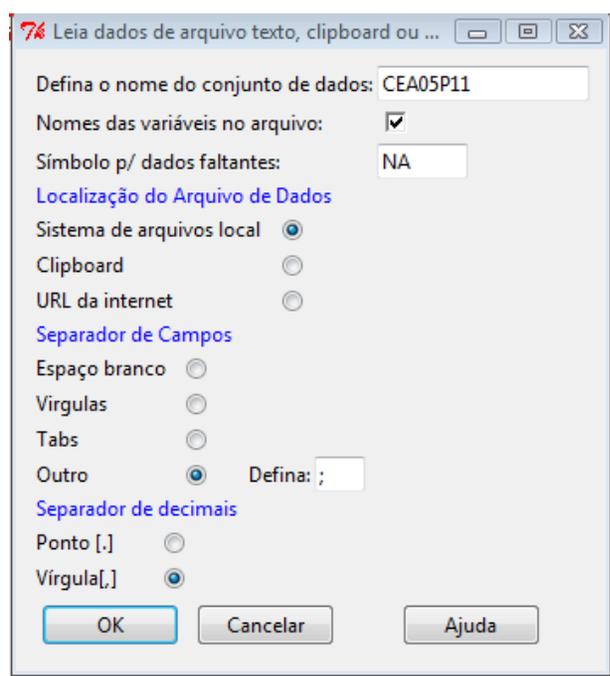
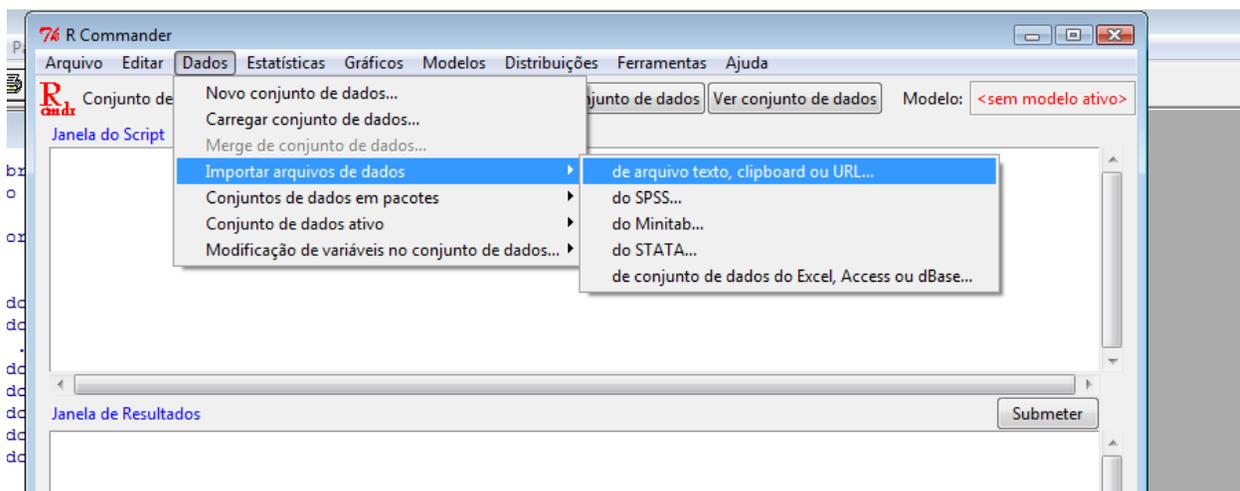
**Vitalidade antes (Vit\_a):** quantitativa contínua

**Estado Geral de Saúde depois (EGS\_d):** quantitativa contínua

**Vitalidade depois (Vit\_d):** quantitativa contínua

(b) **Passo 1:** Lendo os dados do arquivo .csv no Rcmdr:

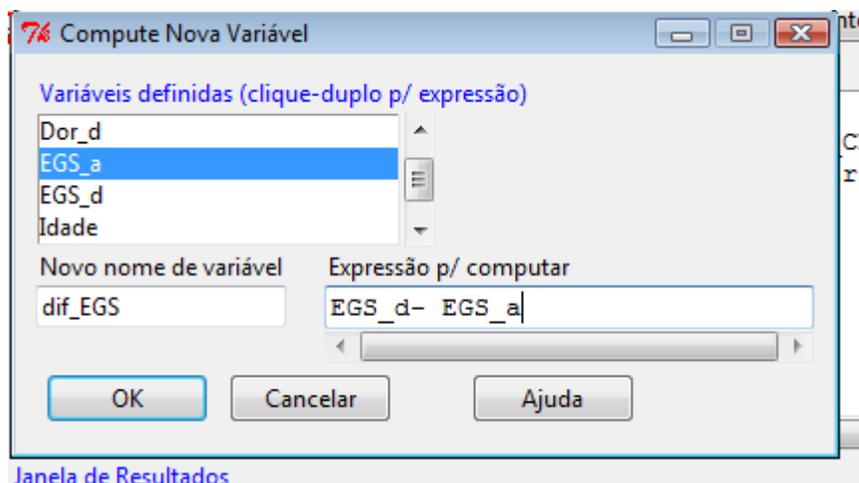
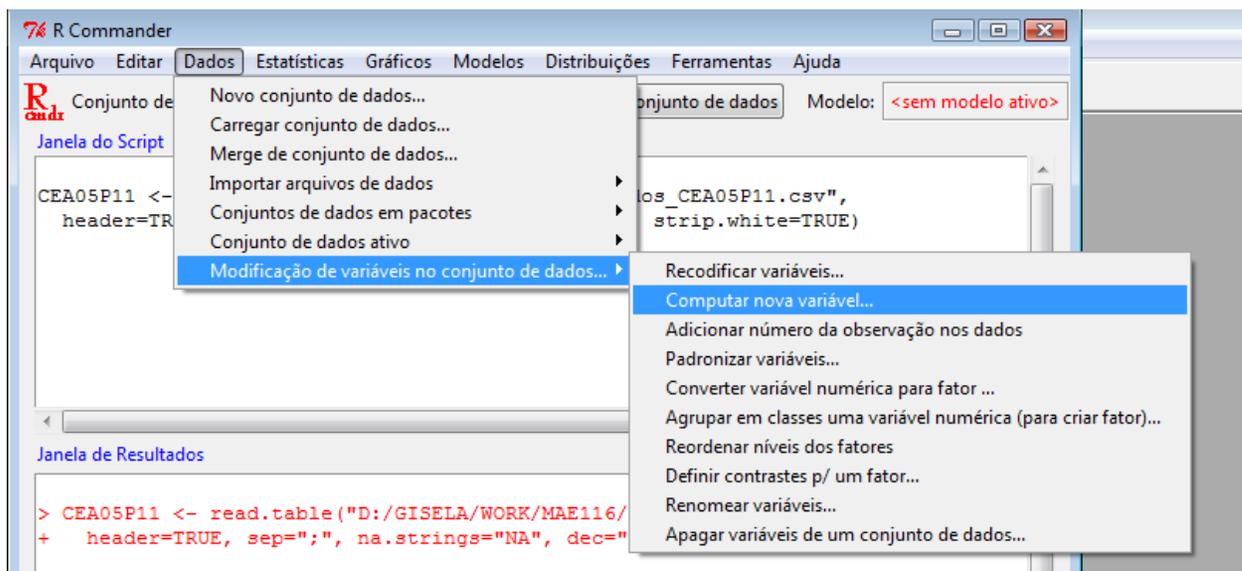
MAE116 – Noções de Estatística  
Grupo A - 1º semestre de 2012 - GABARITO  
Lista de exercícios 2 – Estatística Descritiva – C L A S S E



Seguir instruções

**Passo 2:** Criar a nova variável

MAE116 – Noções de Estatística  
Grupo A - 1º semestre de 2012 - GABARITO  
Lista de exercícios 2 – Estatística Descritiva – C L A S S E



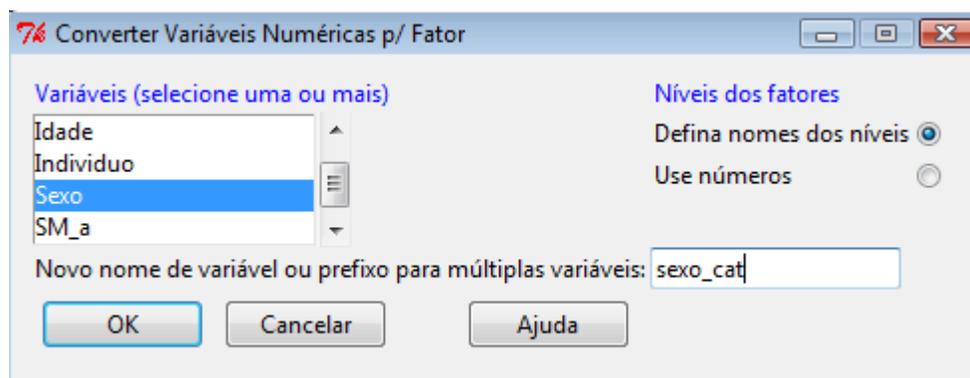
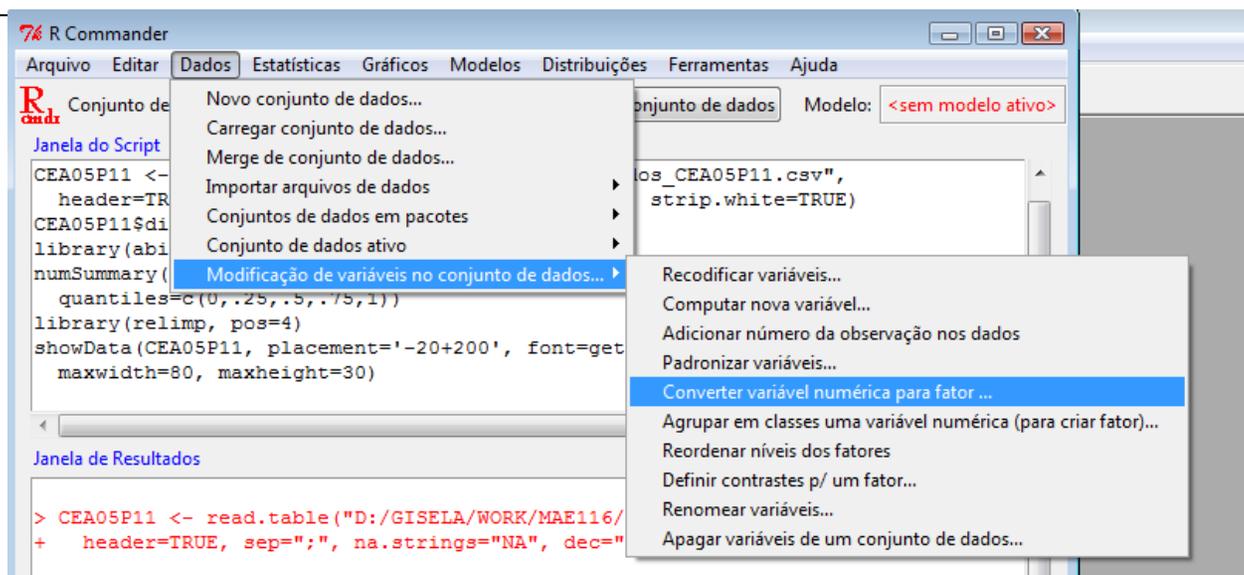
Classificação: variável quantitativa contínua

- (c) **Diferença segundo sexo:** a variável sexo está codificada como 1- Masc e 2 – Fem (ver segunda planilha no arquivo Excel) e o R entende como numérica. É necessário criar uma variável com as categorias:

# MAE116 – Noções de Estatística

## Grupo A - 1º semestre de 2012 - GABARITO

### Lista de exercícios 2 – Estatística Descritiva – C L A S S E



Medidas descritivas:

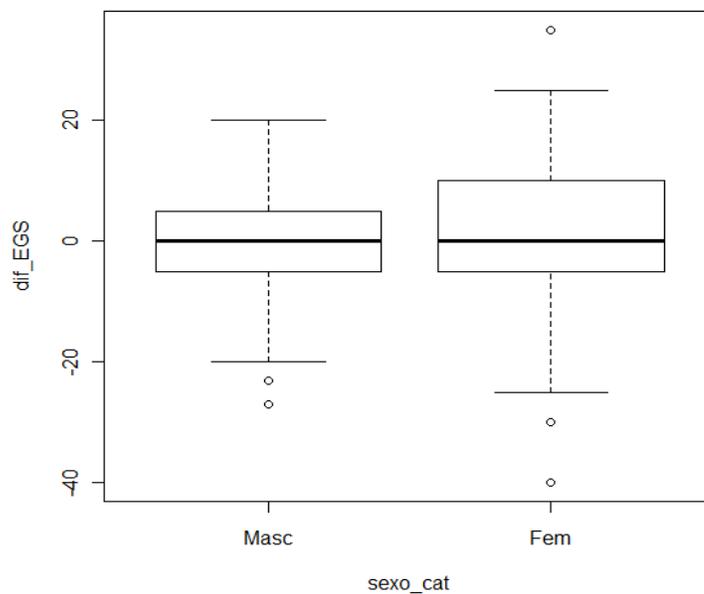
Estatísticas → Resumos → Resumos numéricos → Resuma por grupos

	mean	sd	0%	25%	50%	75%	100%	n
Masc	0.5405405	11.09703	-27	-5	0	5	20	37
Fem	1.4433962	12.59636	-40	-5	0	10	35	106

MAE116 – Noções de Estatística  
Grupo A - 1º semestre de 2012 - GABARITO  
Lista de exercícios 2 – Estatística Descritiva – C L A S S E

---

Box plot



Comentário: variabilidade um pouco maior da melhora para pessoas do sexo feminino. No caso masculino, média e mediana em torno de zero, o que indica que não houve muita melhora após o programa. Em ambos os casos, houve melhora para metade das mulheres e houve piora para a outra metade.